

SEQ ID 1

LOCUS HDU53215 2384 bp DNA linear BCT 15-MAY-1997
 DEFINITION *Haemophilus ducreyi* cytolethal distending toxin protein A (cdtA),
 protein B (cdtB) and protein C (cdtC) genes, complete cds.
 ACCESSION U53215
 VERSION U53215.1 GI:2102681
 KEYWORDS
 SOURCE *Haemophilus ducreyi*
 ORGANISM *Haemophilus ducreyi*
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
 REFERENCE 1 (bases 1 to 2384)
 AUTHORS Cope, L.D., Lumbley, S., Latimer, J.L., Klesney-Tait, J., Stevens, M.K.,
 Johnson, L.S., Purven, M., Munson, R.S. Jr., Lagergard, T., Radolf, J.D.
 and Hansen, E.J.
 TITLE A diffusible cytotoxin of *Haemophilus ducreyi*
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (8), 4056-4061 (1997)
 MEDLINE 97268696
 PUBMED 9108104
 REFERENCE 2 (bases 1 to 2384)
 AUTHORS Hansen, E.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1996) Microbiology, UT Southwestern Medical
 Center, 6000 Harry Hines Blvd NA6.200, Dallas, TX 75235-9048, USA
 FEATURES Location/Qualifiers
 source 1..2384
 /organism="Haemophilus ducreyi"
 /strain="35000"
 /db_xref="taxon:730"
 gene 152..823
 /gene="cdtA"
 CDS 152..823
 /gene="cdtA"
 /codon_start=1
 /transl_table=11
 /product="cytolethal distending toxin protein A"
 /protein_id="AAB57725.1"
 /db_xref="GI:2102682"
 /translation="M KKFLPSLLL MGSVACSSNQ RMNDY SQPESQSDLAPKSSTI QPQ
 PQPLL S KTPSMSL NLLSSSGPNRQVLPSEPSNFMTLMGQ NGALLTVWALAKRNWLWAY
 PNIY SQDFGNIRNWKMEPGK HREYF R FVNQSLGTCV EAGNGLIHDICSLDKLAQEF
 LLPTD SGA VVI KSVS QGRCV TYNPVSTTFY STVTL S VCDGATEPSRDQTWYLAPP
 VLE ATAVN"
 gene 838..1689
 /gene="cdtB"
 CDS 838..1689
 /gene="cdtB"
 /codon_start=1
 /transl_table=11
 /product="cytolethal distending toxin protein B"
 /protein_id="AAB57726.1"
 /db_xref="GI:2102683"
 /translation="M QWVKQLSVVFCVMLFSFSSYANLSDFKVATWNLQGSSAVNESK
 WNI NRQLLSGEQGADILMVQEAGSLPSSAVRTSRVIQHGGTPIE EYTWNLGTRSRPN
 MVYI YYSLR DVGAN RVNL AI VSRR QADEAFIVHSDSSVLQSRPAVGIRIGTDVFTV
 H ALATGGSDAVSLIRNIFT FNSSSPERRVY SWMVVGDFN RAPAN LEVAL RQEPAVS
 ENT IIIIAPTEP THRS GNILDY AILHDA LPR REQARERIGASLMLNQLRSQITS DHFP
 VSFVRDR"
 gene 1700..2260
 /gene="cdtC"
 CDS 1700..2260

```
/gene="cdtC"
/codon_start=1
/transl_table=11
/product="cytotoxic distending toxin protein C"
/protein_id="AAB57727.1"
/db_xref="GI:2102684"
/translation="MKKYLLSFLLIMILALASHAESNPDPPTYPDVELSPPPRISLRS
LLTAQPVKNDHYDSHNYLSTHWELIDYKGKEYEKLRDGGTLVQFKVVGAAKCFAFLGK
GTTDCKDHDHTVFNLIPNTNTGAFLIKDALLGFCITSHDFDDLKLEPCGGSVSGRTFSL
AYOWGILPPFGPSKILIPPVRRNQGS"
```

BASE COUNT 714 a 416 c 494 g 760 t

ORIGIN

11

LOCUS CJU51121 2600 bp DNA linear BCT 28-AUG-1996
 DEFINITION *Campylobacter jejuni* cytolethal distending toxin (*cdtA*), (*cdtB*), (*cdtC*) genes, complete cds, and *LctP* gene, partial cds.
 ACCESSION U51121
 VERSION U51121.1 GI:1354273
 KEYWORDS
 SOURCE *Campylobacter jejuni*
 ORGANISM *Campylobacter jejuni*
 Bacteria; Proteobacteria; epsilon subdivision; *Campylobacter* group;
Campylobacter.
 REFERENCE 1 (bases 1 to 2600)
 AUTHORS Pickett, C.L., Pesci, E.C., Cottle, D.L., Russell, G., Erdem, A.N. and Zeytin, H.
 TITLE Prevalence of cytolethal distending toxin production in *Campylobacter jejuni* and relatedness of *Campylobacter* sp. *cdtB* gene
 JOURNAL *Infect. Immun.* 64 (6), 2070-2078 (1996)
 MEDLINE 96239019
 PUBMED 8675309
 REFERENCE 2 (bases 1 to 2600)
 AUTHORS Pickett, C.L.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAR-1996) Carol L. Pickett, Microbiology/Immunology, University of Kentucky, UKMC 800 Rose St., Lexington, KY 40536, USA
 FEATURES Location/Qualifiers
 source 1..2600
 /organism="Campylobacter jejuni"
 /strain="81-176"
 /db_xref="taxon:197"
 gene 141..947
 /gene="cdtA"
 CDS 141..947
 /gene="cdtA"
 /codon_start=1
 /transl_table=11
 /product="CdtA"
 /protein_id="AAB06707.1"
 /db_xref="GI:1354274"
 /translation="MQKIIIVFILCCFMTFFLYACSSKFENVNPLGRSFGEFEDTDPLK
 LGLEPTFPTNQEIPSLISGADLVPITPITPPLTRTSNSANNNAANGINPRFKDEAFND
 VLIFENRPAVSDFLTILGPSGAALTIVWALAQGNWIWGYTLIDSKGFGDARVWQLLLYP
 NDFAMIKNAKTNCLNAYGNGIVHYPCDASNHAQMWKLIIPMSNTAVQIKNLGNGKCIQ
 APITNLYGDFHKVFKIFTVECAKKDNFDQQWFLTTPPFTAKPLYRQGEVR"
 gene 944..1741
 /gene="cdtB"
 CDS 944..1741
 /gene="cdtB"
 /codon_start=1
 /transl_table=11
 /product="CdtB"
 /protein_id="AAB06708.1"
 /db_xref="GI:1354275"
 /translation="MKKIIICLFLSFNLAFANLENFNVGTWNLQGSSAATESKWSVSVR
 QLVSGANPLDILMIQEAGTLPRTATPTGRHVQQGGTPIDEYEWNLGTLSRPDRVFIYY
 SRVDVGANRVNLAIVSRMQAEEVIVLPPPTTVSRPIIGIRNGNDAFFNIHALANGGTD
 VGAIITAVDAHFANMPQVNWMIAGDFNRPSTITSTVDRELANRIRVVFPTSATQASG
 GTLDYAITGNSNRQQTYTPPLLAAILMLASLRSHIVSDHFPVNFRKF"
 gene 1752..2321
 /gene="cdtC"
 CDS 1752..2321
 /gene="cdtC"

```

/codon_start=1
/transl_table=11
/product="CdtC"
/protein_id="AAB06709.1"
/db_xref="GI:1354276"
/translation="MKKIITLFFMFITLAFATPTGDLKDFTEMVSIRSLETGIFLSAF
RDTSKDPIDQNWNIKEIVL,DELKQKDKLADELPGYVQFTNPKESDLCLAILEDGTF
GAKSCQDDLKDGKLETVFSIMPTTSAVQIRSLVLESDECIVTFFNPNPIQKRGIA
PCTLDPIFFAEVNELMIITPPLTAATPLE"
gene 2502..2600
/gene="LctP"
CDS 2502..>2600
/gene="LctP"
/codon_start=1
/transl_table=11
/protein_id="AAB06710.1"
/db_xref="GI:1511637"
/translation="MLTWQQIYDPFSNIWLSALVAFLPILCFLVCLV"

```

BASE COUNT 873 a 402 c 421 g 904 t

ORIGIN

SEQ ID 2 1 tgctaaaata taagtgttta agatacatat aaattctacc tttaaaaaca acaaaataaa
61 acatttttaa aaagcgaaa attataatga aatttatgtt attattttct taaaaattta
121 aatacatatc aaggtttta atgcaaaaaa ttatagttt tattttatgt tgtttatga
181 cttttttct ttatgcattt tcttctaaat ttgaaaatgt aaatccttg gggcggtcat
241 ttggagaatt tgaagatact gatcctttaa aactaggact tgaacctact ttcctacca
301 atcaagaaaat tccaagtttta attagcggtt ctgatttagt acctattact cctattaccc
361 caccttaaac tagaacaagc aatagtgc acaataatgc agcaaatggg atcaatcctc
421 gcttaaaaga cgaagctttt aatgatgtt taattttga aaatcgccct gcgggttctg
481 atttttaac cattttaggc cctagcgag cagcttaac gttttggct ttagcacaag
541 gaaattggat ttggggctat accttaatcg atagcaaagg atttggcgat gctagagttt
601 ggcaactttt gctttatcct aatgattttg caatgattaa aaatgc当地 accaataactt
661 gtcttaatgc ttatggtaat ggaattgtcc attatccttg tgatgc当地 aatcacgcac
721 aaatgtggaa acttattccct atgagcaata cagcgttca aattttttat tttagaaatg
781 gaaaatgc当地 acaagcacct attacaatc ttatggta ttttc当地 aag gttttttaaa
841 ttttaccgt agagtgtgc当地 aaaaagata attttgatca acaatggtt ttaactactc
901 cacctttac cgcaaaaacct ttatatgc当地 aaggagaggt acgatgaaaa aaatttatg
961 tttattttta tcttttaacc ttgcttttgc aaatttagaa aattttatg ttggcacttg
1021 gaatttgcaa ggctcatccg cagccacaga aagcaatgg agtgttagt taagacaact
1081 tgtaagtgg acaaaccct tagatatctt aatgatacaaa gaagcagggaa ctttaccaag
1141 aacagccact ccaacaggac gccatgtgc当地 acaaggtggaa acacctattt atgaatatg
1201 gtggaattta ggaactctt caaggcctga tagggtttt atttattt ctcgc当地
1261 tgttaggagct aatcggttaa atttagctt agtttcaaga atgcaagctg aagaagtgtat
1321 tgtttacct ccacctacta cagttcaag acccattata ggaattcgca atggaaatg
1381 tgcttttttcc aatatccatg cttagctaa tggaggaaca gatgttaggag caattatcac
1441 agctgttagat gcacattttg caaatatgcc tcaagtttac tggatgatag caggggattt
1501 taaccgtgat ctttctacta taacaagtac agtggataga gaatttagcaa atagaattag
1561 agtggttttt ccaactagcg caactcaagc aagcggaggg actcttgatt atgcaattac
1621 agggaaattca aatagacaac aaacctatac tccaccgctt ttagctgc当地 ttttaatgct
1681 tgcaagtttta agatctcata tagtttccaa tcattttccaa gtaaattttt gaaaattttt
1741 ggacatttaa tatgaaaaaa attattactt tgttttttat gtttataact ttgc当地
1801 caactcctac tggagatttgg aaagattttt ccgaaatggg ttctataaga agcttagaaa
1861 cggaaatttt ttaagcgcc tttagggata cctcaaaaaga tcctattgtt caaaattgg
1921 atattaaaga aattgtttta agcgatgagt taaaacaaaa agataaatta gctgatgaac
1981 ttcctttgg ttatgtgcaa ttacaatc caaaagaaaag cgatcttgc当地 ttagccatct
2041 tagaagatgg aacctttgg acaaattttt gtcaagatga tctaaaagat ggtaaattag
2101 aaactgtatt ttctataatg ccaacaacaa cttcagctgt gcaaattcgt tctttatgtt
2161 tggaatctga tgaatgtata gtaactttt ttaatccaaa tattcctata caaaaacgct
2221 ttggatagc cccttgcacc ctagatccta tttttttgc tgaagtaat gaactaatg
2281 ttataacccc acctttaaaca gctgctaccc ctttagaata agattttt cttgttctat
2341 ttttatattt atttaatatt tatgatatta ctaaaataca caaaataatt aataataata

2401 caatgtatt taccttgctc tataattttt ttatTTTaat gtaattttt gttacaataa
2461 atttatacat aataattatc ttggaggaaa aattggaaaca aattttaaaca tggcaacaaa
2521 tttatgaccc ttttcaaAT attggctaa gtgcttagt ggcattttt cctatactat
2581 gtttttagt ttgtttggtt

//

SEQ D 3

LOCUS ECCDTABC 2305 bp DNA linear BCT 29-JAN-1999
 DEFINITION Escherichia coli E6468/62 (O86:H34) cytolethal distending toxin
 (cdtA, cdtB, cdtC) genes, complete cds.
 ACCESSION U03293
 VERSION U03293.1 GI:416213
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 2305)
 AUTHORS Scott, D.A. and Kaper, J.B.
 TITLE Cloning and sequencing of the genes encoding Escherichia coli
 cytolethal distending toxin
 JOURNAL Infect. Immun. 62 (1), 244-251 (1994)
 MEDLINE 94086109
 PUBMED 8262635
 REFERENCE 2 (bases 1 to 2305)
 AUTHORS Scott, D.A.
 TITLE Direct Submission
 JOURNAL Submitted (09-NOV-1993) Scott D.A., University of Maryland School
 of Medicine, Geographic Med., Center for Vaccine Dev., 10 South
 Pine Street, Baltimore, MD 21201, USA
 FEATURES Location/Qualifiers
 source 1..2305
 /organism="Escherichia coli"
 /strain="E6468/62 (O86:H34)"
 /specific_host="Homo sapiens"
 /db_xref="taxon:562"
 /clone="pDS7.96"
 gene 201..914
 /gene="cdtA"
 CDS 201..914
 /gene="cdtA"
 /note="GTG start codon"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CdtA"
 /protein_id="AAD10621.1"
 /db_xref="GI:416214"
 /translation="MDKKLIAFLCTLITGCSNGIGDSPSPGKVELVGIPGQGIAV
 TSNGATPTLGANNTTEFPEVSIMSTGGALLTIWARPVRNWLWGYTPFDNSVNFGENRNWK
 VVDGKDAGTVKFVNVAQGTCMEAFKNGVIHNTCDDNSLSQEFQLLPSTNGNVLIRSSA
 LQTCIRADYLSRTILSPFAFTITLEKCPGAKEETQEMLWAISPPVRAAKPNLIKPELR
 PFRPLPIPDKPDGMEGV"
 gene 911..1732
 /gene="cdtB"
 CDS 911..1732
 /gene="cdtB"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CdtB"
 /protein_id="AAD10622.1"
 /db_xref="GI:416215"
 /translation="MKKLLFLLMILPGISFADLSDFKVATWNLQGSNAPTENKWNTHV
 RQLVTGSGAVDILMVQEAGAVPASATLTEREFSTPGIPMNEYIWNTGTNSRPQELFIY
 FSRVDAFANRVNLAIIVSNRADEVIVLPPPTVSRPIIGIRIGNDVFFSTHALANRGV
 DSGAIVNSVFEFFNRQTDPIRQAANWMIAGDFNRSPATLFSTLEPGIRNHVNIIAPPD

PTQASGGVLDYAVVGNCSVFVLPLLARASLLFGLLRGQIASDHFPVGFIPGRGARR"
 gene 1729..2301
 /gene="cdtC"
 CDS 1729..2301
 /gene="cdtC"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CdtC"
 /protein_id="AAD10623.1"
 /db_xref="GI:416216"
 /translation="MKTIVVFFVLLLTGCASEPANQRNLLTQFVGNNAPVDPEPSPV
 VNIRNVLTGGIIRNPVGSDFNVNNWVISEVKTNLDLISAPGGHVQIKNPDGNECF
 AI LNGQLAVAKQCSESDRNALFTFITSDTGAVQIKSIGSGQCLGN
 GESITDFRLKKCDD LGRPFDTVPPGLLWMLNPLSPA
 MSPLTS"

BASE COUNT 641 a 415 c 547 g 702 t

ORIGIN

SEQ ID 3
 1 tattgaatag ttttgggggg aatataaaaga attatatttgc agtatgctgt ttgtgactct
 61 gggaaaataa acgagcatta taataaaactt gtttgtttt ctggtttgcg atttcctcat
 121 taatcttgtt ggtaatattt tcgttgctt gttgttcta ttttttata aagaagaggt
 181 ggtcagagg agggaaataca gtggataaaa aactaattgc atttttgtgc acacttataa
 241 ttactgggtt ctcgaatggg atcgggtgatt cacccacc tccggggaaa aatgtagaat
 301 tggggaaat ccctggacaa ggtattgcag tgacttcaaa cggtgcaact ccaacacttg
 361 gagccaaacaa cactgagttt cctgaagttt caataatgag cactgggtgg ggcgtgc
 421 ctatgggc cagacctgtt cgttaactggc tttggggta tactcctttt gattcagtaa
 481 attttggta gaatcggaaac tggaaagggtt tggatggaa agatgccggc acagtggaaat
 541 ttgttaatgt tgcccagggg acttgcatttgg aggcccttaa aaacgggggtg atacataata
 601 cctgtatgtt taactcgat ttcaggat ttcaggat ttcaggat ttcaggat ttcaggat
 661 tgcttataag aagtagtgcc ttgcagacgt gtataagagc agactattta agcagaacta
 721 tattgtcacc gtttgc tttt acaatcacc tttggaaaatg ccctgggtca aaagaagaaaa
 781 cgcaagaaaat gctatggca ataagtccac ctgtcagagc ggcaaaaacca aatctgatta
 841 agccagagtt aagaccatttcc agaccatttgc caattccacc tcatgacaaa cctgatggaa
 901 tggagggagt atgaaaaaaat tattattcctt gttaatgatt ttgcgggta tttctttgc
 961 agatttaagc gattttaaag ttgcacccctt gattttgcag gttcaatgc caccgacaga
 1021 aaataaaatgg aacacacatg tccgacaact ttttgcgggaa atgtgggtctg ttgtatcc
 1081 gatggttcag gaggcagggg cagtaccatg ttctgcacatg ttgtactgagc gagaatttag
 1141 cactccttgtt attcgcatttgc atgatgtat tttttttttt gttttttttt tttttttttt
 1201 ggagttgtt atatatttgc cacgtgttgc tgcatttcgc aacagatgaa atcttgc
 1261 tggggaaatgg agaagagctg atgagggtat tttttttttt cttttttttt tttttttttt
 1321 accgatcatc ggcattagaa ttggtaatgtt tttttttttt tttttttttt tttttttttt
 1381 tcggggcgtt gatttcaggat gttttttttt tttttttttt tttttttttt tttttttttt
 1441 ggatcctata agacaggccg ctaactggat gttttttttt tttttttttt tttttttttt
 1501 tacactattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1561 agatccaacg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1621 ctttgcattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1681 ctctgatcat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1741 atttgttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1801 ctcttactt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1861 gttaatatca gaaacgttct tttttttttt tttttttttt tttttttttt tttttttttt
 1921 aatgttaataa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 1981 ccggggaggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 2041 gggcaattgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 2101 ataaccatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 2161 aatggagaga tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 2221 ttttgcattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 2281 atgtctccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt

//

LOCUS ECU04208 2600 bp DNA linear BCT 07-JUN-1994
 DEFINITION Escherichia coli 9142-88 cytolethal distending toxin (cdtA, cdtB, and cdtC) genes, complete cds.
 ACCESSION U04208
 VERSION U04208.1 GI:436944
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 2600)
 AUTHORS Pickett,C.L., Cottle,D.L., Pesci,E.C. and Bikah,G.
 TITLE Cloning, sequencing, and expression of the Escherichia coli cytolethal distending toxin genes
 JOURNAL Infect. Immun. 62 (3), 1046-1051 (1994)
 MEDLINE 94156453
 PUBMED 8112838
 REFERENCE 2 (bases 1 to 2600)
 AUTHORS Pickett,C.L.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1993) C.L. Pickett, University of Kentucky, Microbiology and Immunology, Dept. of Micro and Immuno - Chandler Medical Center, Lexington, KY 40536, USA
 FEATURES Location/Qualifiers
 source 1..2600
 /organism="Escherichia coli"
 /strain="9142-88"
 /db_xref="taxon:562"
 gene 408..1184
 /gene="cdtA"
 CDS 408..1184
 /gene="cdtA"
 /codon_start=1
 /transl_table=11
 /product="CdtA"
 /protein_id="AAA18785.1"
 /db_xref="GI:436945"
 /translation="MANKRTPIFIAGILIPILLNGCSSGKNKAYLDPKVFPQVEGGP
 TVPSPDEPGLPLPGPGPALPTNGAIPAPIPEPGTAPAVSLMNMDGSVLTWSRGAGSSLW
 AYYIGDSNSFGELRNWQIMPQGTRPNTIQFRNVDVGTCTMSFPGFKGGVQLSTAPCKFG
 PERFDQPMATRNGNYQLKSLSTGLCIRANFLGRTPSSPYATTLMERCPSSEKNFE
 FMWSISEPLRPALATIAKPEIRPFPQPIEPDEHSTGGEQ"
 gene 1181..1990
 /gene="cdtB"
 CDS 1181..1990
 /gene="cdtB"
 /codon_start=1
 /transl_table=11
 /product="CdtB"
 /protein_id="AAA18786.1"
 /db_xref="GI:436946"
 /translation="MKKYIISLIVFLSFYAQADLTDFRVATWNLQGASATTESKWNIN
 VRQLISGENAVDILAVQEAGSPPSTAVDTGTLIPSPGIPVRELIWNLSTNSRPQQVYI
 YFSAVDALGGRVNLALVSNRRADEVFVLSPVRQGGRPLLGIIGNDAFTAHAIAMRN
 NDAPALVEEVYNFFRDSRDPVHQALNWMLGDFNREPADLEMLTVPVRRASEIISPA
 AATQTSQRTLDYAVAGNSVAFRPSPLQAGIVYGARRTQISSDHFPVGVSRR"
 gene 2005..2550
 /gene="cdtC"
 CDS 2005..2550
 /gene="cdtC"

```

/codon_start=1
/transl_table=11
/product="CdtC"
/protein_id="AAA18787.1"
/db_xref="GI:436947"
/translation="MKKLAIVFTMLLIAGCSSSQDSANNQIDELGKENNSLFTFRNIQ
SGLMIHNGLHQHGRETIWVEIPVKTPEEALVTDQSGWIMIRTPNTDQCLGTPDGRNL
LKMTCNSTAKTTLFSLIPSTTGAVQIKSVLSGLCFLDSKNGLSFETGKCIADFKKP
EVVPQSHLWMLNPLNTESPII"

```

BASE COUNT 834 a 555 c 508 g 703 t
 ORIGIN

SEQ ID 4

```

1 attaaacaaat tacaacaaag atcacattaa ataaaaattg acaattgacc tgtagttcat
61 catttgtaaa ttcatgattt atatcaatca cgctttgtgt tcggagtaag cttataaatt
121 acaaaaacga ttaaataaaaa aaccacacaa taatattaaa taaaataacg gtatcgactg
181 cttttgttcc aaaaggaatt gctttaaaaa ctatatactt tcatttagtt ttatcaatta
241 atgcataact tcaaattgtaa catcaaaaac aatacacaccc aaaacaatca cacaagcaa
301 caaggacacc caaacaacta aggactaat aaaaaggaga gtcccaatgt aattttttt
361 ttcttccatt aatttctact atctttatca taataaggac cataataatg gctaacaac
421 gtacacccat ttttatactt ggaatcttga tccccattttt attaaatggg tgctcatcag
481 gaaaaaaataa agcttatctt gaccccaaag ttttcctcc tcaagtggaa ggaggaccaa
541 ccgttccttc ccccgatgag cccggacttc cattggccgg gccaggaccc gcgctgccc
601 caaatggcgc aatccctatc cctgaaccag gtacccgacc cgcagatctt ttaatgaata
661 tggatggctc agttctaaaca atgtggagcc gcgagctgg ttcatcgta tggcgtatt
721 atatcgccgaa ctccaaattca tttggggaaac tacgttaattt gcagattatg cccgaaacca
781 ggc当地atac gatacagttt cgc当地atgtac acgttggtag ctgtatgaca agttccac
841 gatttaaagg gggagttacaa ctttctacag caccttgcaa gtttggaccc gaacgtttcg
901 atttccagcc aatggcaaca cgc当地atggta attaccagtt aaaaatctt tctacaggtt
961 tatgcatcag agc当地atttt ttaggaagaa caccatcatc tccgtacgc acgacattaa
1021 caatggagcg ttggccatca agtggagaga aaaaacttta attcatgtgg tccataagcg
1081 aaccattaaag gcctgcctcg gcaactattt ccaaggccaga aatacgccca tttccaccac
1141 agccaaataga accagatgag cattcaactg gaggagaaca atgaaaaaaat atattatatc
1201 tctgatagtg tttttatcat ttacgctca agcagatttta actgattttc gcggtgcac
1261 ctggaaatctt caaggtgcat ccgc当地acac tgaaagtaaa tggaaatataa atgtccggca
1321 attaatttctt ggtgaaaatg ctgttagacat tttagtgc当地tta caagaggcag gctccgc
1381 gtcaacggct gtagatacag gtacacttat tccttccca ggaattcccg tccgagagct
1441 tatctggAAC ttgtcgacaa atagcaggcc acagcaagta tatatatatt tttccgctgt
1501 tgatggccctc ggtggaaagag tcaatcttgc tctggtagc aatcgccggg cc当地atgaa
1561 gtttggctt agtcctgtaa gacaagggtgg acgaccatttgc ttggcatac gaattggtaa
1621 tgatgcattt ttcaactgc当地 acgc当地atgc tatgc当地aaac aatgatgccc cggcttgc
1681 tgaggaagtg tataacttctt tccgc当地acag cagagacccca gtacaccagg cgcttaactg
1741 gatgattctt ggtgatttca accgtgaaacc tgc当地gattttta gagatgaaacc ttactgttcc
1801 cgtaaagg gcatc当地aaa ttatccacc agcggccggca acacaaacca gccagcgaac
1861 attagattat gcaatgc当地 gaaactctgt ggc当地ttaga ccatctccgc tacaagcggg
1921 aattgtatatt ggagccagga gaactcaat atcttgc当地 catttccctg tggcgtatc
1981 cagacgataa aagaggctat cataatgaaa aaattagcaa ttgtttttac tatgctgcta
2041 atagctggat gctcttcatc acaggattca gctaacaatc agatagatga attaggaaaa
2101 gaaaacaatt ctctattcatc attccgc当地 atccaaatgt gcttaatgtat ccataatgga
2161 ttacatcagc atggccgaga gactatttggg tggaaatag tccctgtgaa aacacctgaa
2221 gaagcacttg ttaccgatca aagcgggtgg ataatgatttgc gaaacccaaa cacagaccaa
2281 tggttaggaa cgcctgatgg aaggaaacccctt ctaaaaaatgtatc gctgtatc aacagctaa
2341 aaaacttgc当地 tttctctcat accgtcaaca acaggggcag tacaatcaa aagcgttctg
2401 tctggccctt gtttcttgc当地 tagttaaaaat agcggattaa gtttggaaac gggaaatgc
2461 attgctgact tcaaaaaacc atttgaatgtt gtagccacaga gccatttgc当地 gatgttgaac
2521 ccattaaata ctgaatgc当地 tattatatttgc当地 tcccatcatc gcatggcc gggcacataa
2581 aaagcattat cataataatgt
  //

```

SEQ ID 5

LOCUS AF373206 468 bp DNA linear BCT 02-MAY-2002
DEFINITION Escherichia coli strain MBU. E 412 cytolethal distending toxin B
(cdtB) gene, partial cds.
ACCESSION AF373206
VERSION AF373206.1 GI:20385547
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 468)
AUTHORS Bouzari, S., Oloomi, M. and Zarepoor, M.
TITLE Identification of cdtB homologs in diarrheagenic E. coli isolates in
Iran
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 468)
AUTHORS Bouzari, S., Oloomi, M. and Zarepoor, M.
TITLE Direct Submission
JOURNAL Submitted (22-APR-2001) Molecular Biology, Institute Pasteur of
Iran, Pasteur Ave., Tehran 13164, Iran
FEATURES Location/Qualifiers
source 1..468
/organism="Escherichia coli"
/strain="MBU. E 412"
/db_xref="taxon:562"
gene <1..>468
/gene="cdtB"
CDS <1..>468
/gene="cdtB"
/note="CdtB"
/codon_start=2
/transl_table=11
/product="cytolethal distending toxin B"
/protein_id="AAM21324.1"
/db_xref="GI:20385548"
/translation="ENKWNTHVRQLVTGSGAVDILMVQEAGAVPASATLTEREFSTPG
IPMNEYIWNTGTNSRPQELFIYFSRVDAFANRVNLAIIVSNRADEVIVLPPPTVSRP
IIGIRIGNDVFFSTHALANRGVDSGAIVNSVFEFFNRQTDPIRQAANWMIAGD"
BASE COUNT 130 a 89 c 115 g 134 t
ORIGIN

1 tgaaaataaaa tggaacacac atgtccgaca acttgttacg ggaagtgggt ctgttgatat
61 cctgatggtt caggaggcag gggcagtacc agcttctgca acgttgactg agcgagaatt
121 tagcaactcct ggtattccga tgaatgagta tatctggaat accggAACCA atagtcgccc
181 acaggagttg tttatataatt ttcacgtgt tgatgcattc gctaacagag taaaatcttgc
241 gattgtttca aacagaagag ctgatgaggt gattgttata cctccctccaa ctgttgatc
301 acgaccgatc atcggcatttta gaattggtaa tgatgttttc ttctcaaccc atgcattggc
361 gaatcggggc gtggattcag gagcaattgt aaatagtgtt ttgagttct tcaacagaca
421 aacggatcct ataagacagg ccgcttaactg gatgattgca ggagattt

//

S6 106

LOCUS G22P1 **2743 bp** **mRNA** **linear** **PRI** **05-NOV-2002**
DEFINITION Homo sapiens thyroid autoantigen 70kDa (Ku antigen) (G22P1), mRNA.
ACCESSION NM_001469
VERSION NM_001469.2 GI:20070134
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2743)
AUTHORS Chan, J.Y., Lerman, M.I., Prabhakar, B.S., Isozaki, O., Santisteban, P.,
Kuppers, R.C., Oates, E.L., Notkins, A.L. and Kohn, L.D.
TITLE Cloning and characterization of a cDNA that encodes a 70-kDa novel
human thyroid autoantigen
JOURNAL J. Biol. Chem. 264 (7), 3651-3654 (1989)
MEDLINE 89139411
PUBMED 2917966
REFERENCE 2 (bases 1 to 2743)
AUTHORS Reeves, W.H. and Sthoeger, Z.M.
TITLE Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen
JOURNAL J. Biol. Chem. 264 (9), 5047-5052 (1989)
MEDLINE 89174787
PUBMED 2466842
REFERENCE 3 (bases 1 to 2743)
AUTHORS Griffith, A.J., Craft, J., Evans, J., Mimori, T. and Hardin, J.A.
TITLE Nucleotide sequence and genomic structure analyses of the p70
subunit of the human Ku autoantigen: evidence for a family of genes
encoding Ku (p70)-related polypeptides
JOURNAL Mol. Biol. Rep. 16 (2), 91-97 (1992)
MEDLINE 92301477
PUBMED 1608402
REFERENCE 4 (bases 1 to 2743)
AUTHORS Tuteja, N., Tuteja, R., Ochem, A., Taneja, P., Huang, N.W.,
Simoncsits, A., Susic, S., Rahman, K., Marusic, L., Chen, J. et al.
TITLE Human DNA helicase II: a novel DNA unwinding enzyme identified as
the Ku autoantigen
JOURNAL EMBO J. 13 (20), 4991-5001 (1994)
MEDLINE 95045391
PUBMED 7957065
REFERENCE 5 (bases 1 to 2743)
AUTHORS Koike, M., Matsuda, Y., Mimori, T., Harada, Y.N., Shiomi, N. and
Shiomi, T.
TITLE Chromosomal localization of the mouse and rat DNA double-strand
break repair genes Ku p70 and Ku p80/XRCC5 and their mRNA
expression in various mouse tissues
JOURNAL Genomics 38 (1), 38-44 (1996)
MEDLINE 97124844
PUBMED 8954777
REFERENCE 6 (bases 1 to 2743)
AUTHORS Baumann, P. and West, S.C.
TITLE DNA end-joining catalyzed by human cell-free extracts
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (24), 14066-14070 (1998)
MEDLINE 99045638
PUBMED 9826654
REFERENCE 7 (bases 1 to 2743)
AUTHORS Goedcke, W., Eijpe, M., Offenberg, H.H., van Aalderen, M. and
Heyting, C.
TITLE Mre11 and Ku70 interact in somatic cells, but are differentially
expressed in early meiosis
JOURNAL Nat. Genet. 23 (2), 194-198 (1999)

MEDLINE 99438394
 PUBMED 10508516
 COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from BC008343.1. On Apr 8, 2002 this sequence version replaced gi:4503840.
 Summary: The G22P1 gene encodes subunit p70 of the p70/p80 autoantigen. The p70/p80 autoantigen consists of 2 proteins of molecular mass of approximately 70,000 and 80,000 daltons that dimerize to form a 10 S DNA-binding complex. See MIM 194364 for discussion of the gene encoding the p80 subunit. Exchange of immunologic reagents showed that the p70/p80 autoantigen is identical to the Ku antigen, the Ki antigen, and the 86- to 70-kD protein complex. The p70/p80 complex binds to the ends of double-stranded DNA in a cell cycle-dependent manner, being associated with chromosomes of interphase cells, followed by complete dissociation from the condensing chromosomes in early prophase. Both p70 and p80 contain phosphoserine residues. A role for the antigen in DNA repair or transposition has been proposed. [supplied by OMIM].
 FEATURES Location/Qualifiers
 source 1..2743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="22q13.2-q13.31"
 /clone="MGC:15860 IMAGE:3509801"
 /tissue_type="Brain, neuroblastoma"
 /clone_lib="NIH_MGC_19"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 gene 1..2743
 /gene="G22P1"
 /note="synonyms: ML8, KU70, TLAA, D22S671, D22S731"
 /db_xref="LocusID:2547"
 /db_xref="MIM:152690"
 CDS 656..2485
 /gene="G22P1"
 /note="thyroid autoantigen 70kD (Ku antigen)"
 /codon_start=1
 /product="thyroid autoantigen 70kDa (Ku antigen)"
 /protein_id="NP_001460.1"
 /db_xref="GI:4503841"
 /db_xref="LocusID:2547"
 /db_xref="MIM:152690"
 /translation="MSGWESYYKTEGDEEAEEEQEEENLEASGDYKYSGRDSLIFLVDA
 SKAMFESQSEDELTPFDMSIQCISQVYISKIISSSDRLLAVVYGTGTEKDKNFVNFKN
 YVLQELDNPAGRAKILELDQFKGQQGQKRFQDMMGHGSDYSLSEVLWVCANLFSDVQFK
 MSHKRIMLFTNEDNPNGNDSAKASRARTKAGDLRDTGIFLFLDMHLKKPGGFISLIFYR
 DIISIAEDEDLVRVHFEESSKLEDLLRKVRAKETRKRALSRKLKLNKDIVISVGIYNL
 VQKALKPPPPIKLYRETNEPVKTKRTFTNSTGGLLPSDTKRSQIYGSRQIILEKEET
 EELKRFDDPGLMLMGFKPLVLLKKHHYLRPSLFVYPEESLVIGSSTLFSALLIKCLEK
 EVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPPGFQLVFLPFADDKRKMPFTE
 KIMATPEQVGKMKAIIVEKLRFTYRSDSFENPVLQQHFRNLEALALDLMEPEQAVDLTL
 PKVEAMNKRLGSLVDEFKELVYPPDYNPEGKVTKRKHNEGSGSKRPKVEYSEEELKT
 HISKGTGLKFTVPMLEACRAYGLKSGLKQELLEALTKHFQD"
 misc_feature 764..2314
 /gene="G22P1"
 /note="ku; Region: A family of DNA binding domains including those in Ku70/80"
 /db_xref="CDD:LOAD_ku"

```

misc_feature 764..1408
/gene="G22P1"
/note="Ku_N; Region: Ku70/Ku80 N-terminal alpha/beta
domain. The Ku heterodimer (composed of Ku70 and Ku80)
contributes to genomic integrity through its ability to
bind DNA double-strand breaks and facilitate repair by the
non-homologous end-joining pathway. This is the amino
terminal alpha/beta domain. This domain only makes a small
contribution to the dimer interface. The domain comprises
a six stranded beta sheet of the Rossman fold"
/db_xref="CDD:pfam03731"
1451..2059
/gene="G22P1"
/note="Ku; Region: Ku70/Ku80 beta-barrel domain. The Ku
heterodimer (composed of Ku70 and Ku80) contributes to
genomic integrity through its ability to bind DNA
double-strand breaks and facilitate repair by the
non-homologous end-joining pathway. This is the central
DNA-binding beta-barrel domain. This domain is found in
both the Ku70 and Ku80 proteins that form a DNA binding
heterodimer"
/db_xref="CDD:pfam02735"
1577..2017
/gene="G22P1"
/note="Ku78; Region: Ku70 and Ku80 are 70kDa and 80kDa
subunits of the Lupus Ku autoantigen"
/db_xref="CDD:smart00559"
2066..2332
/gene="G22P1"
/note="Ku_C; Region: Ku70/Ku80 C-terminal arm. The Ku
heterodimer (composed of Ku70 and Ku80) contributes to
genomic integrity through its ability to bind DNA
double-strand breaks and facilitate repair by the
non-homologous end-joining pathway. This is the C terminal
arm. This alpha helical region embraces the beta-barrel
domain pfam02735 of the opposite subunit"
/db_xref="CDD:pfam03730"
2372..2476
/gene="G22P1"
/note="SAP; Region: SAP domain. The SAP (after SAF-A/B,
Acinus and PIAS) motif is a putative DNA binding domain
found in diverse nuclear proteins"
/db_xref="CDD:pfam02037"
2372..2473
/gene="G22P1"
/note="SAP; Region: Putative DNA-binding (bihelical) motif
predicted to be involved in chromosomal organisation"
/db_xref="CDD:smart00513"
variation 1472
/gene="G22P1"
/allele="A"
/allele="G"
/db_xref="dbSNP:2229034"
variation 2256
/gene="G22P1"
/note="WARNING: map location ambiguous"
/allele="A"
/allele="T"
/db_xref="dbSNP:1803107"
variation 2402

```

/gene="G22P1"
 /allele="A"
 /allele="G"
 /db_xref="dbSNP:3179615"
 variation 2434
 /gene="G22P1"
 /allele="G"
 /allele="T"
 /db_xref="dbSNP:132788"
 BASE COUNT 732 a 629 c 743 g 639 t
 ORIGIN
 SEQ ID 6

```

1 gcggggccgtt atccatttgc gttgttcgcc agctaggcct ggcctcgcc cgcttcgctc
61 ggtcggcttc gcgcgcggcc atagccttgc tagagggtta gcgttagcct taagtgtgcg
121 aatccgagga gcagcgcacag actcgagacc acgctcccttc ctcggaaagg aggccggcacc
181 tcgcgttgc ggcccgccctg cgtttgaggc cccgcgtgc ttgcggcccg cctgcgttg
241 aggccctgtct gcgtttgaga ttcattttgg cgtgattgag gaattttggg aggtttttgg
301 gcggtattga ggacgagggg gtccgttagt cagcatagaa tcctggagcg ggaatccctc
361 accgtctaaa tggcgctcggg ggcgggaccc cccggatctg gcttccgcgg gccggcccg
421 gcccgtaaac gtgaggggata gctgagatga ggcagactact gggatggccc ccatgcgcatt
481 ttacatgcag tccgactgcc gagcttgcg ggcagcagga ttaccgtcc acattccctca
541 ctactaaccat agcttttgcg acagatctca caagaaccta gaggtcggtt tttttcgat
601 ttaaatttgc ctgttactga ctgttaacgtc tttccgttag tgagcagtag ccaacatgtc
661 aggggtggag tcatattaca aaaccgaggg cgtatggaa gcaaggaa aacaagaaga
721 gaaccttgcgaa gcaagtggag actataaata ttcaggaa gatagttga tttttttgg
781 ttagtgcctcc aaggctatgt ttgaatctca gagtgaagat gagttgacac cttttgacat
841 gagcatccag tggatccaaa gttgttgcgtat cagtaagatc ataaggcgtg atcgagatct
901 cttggctgtg gtgttctatg gtaccggaaa agacaaaaat tcaatggaaatt ttaaaaat
961 ttacgtctta caggagctgg ataatccagg tgcaaaacga attcttagac ttgaccagtt
1021 taaggggcag cagggacaaa aacgtttcca agacatgtg gccacggat ctgactactc
1081 actcgtgaa gtgtgtggg tctgtgcctt cctctttgtat gatgtccat tcaagatgag
1141 tcataagagg atcatgtgt tcaccaatga agacaacccc catggcaatg acatgtccaa
1201 agccagccgg gcccggacca aagccgggtga tctccggat acaggcatct tccttgactt
1261 gatgcacccgt aagaaacctg ggggcttgcat cttccgttgc ttctacagag atatcatcag
1321 catagcagag gatgaggacc tcagggttca ctttggggaa tccagcaagc tagaagaccc
1381 gttcgcccaag gttcgccca aggagaccag gaagcggca ctcaggatgt taaagctgaa
1441 gctcaacaaa gatatagtga tctctgtggg catttataat ctggtccaga aggctctcaa
1501 gcctccttcca ataaagctct atcgggaaac aaatgaacca gtgaaaacca agacccggac
1561 cttaatacata agtacaggcg gtttgcctt gcctacgtat accaagaggt ctcagatcta
1621 tggagactgtt cagattatac tggagaaaga gggaaacagaa gagctaaaac gtttgcgt
1681 tccagggttgc atgctcatgg gtttcaagcc gttgttactg ctgaagaaa accattaccc
1741 gagggccctcc ctgttgcgtt acccagagga gtcgtgggtt attggggactt caaccctgtt
1801 cagtgcctgt ctcatcaagt gtctggagaa ggagggttgc gatgtgcata gatacacacc
1861 cccgaggaaac atccctccctt atttgtggc tttgggttgc caggaagaag agttggatga
1921 ccagaaaaatt caggtgactc ctccaggctt ccagctggc tttttacccct ttgtgtatga
1981 taaaagggaaat atgccttta ctgaaaaat catggcaact ccagagcagg tggcaagat
2041 gaaggctatc gttgagaagc ttgcgttccat atacagaagt gacagcttgc agaaccggcgt
2101 gctgcagcag cacttcagga acctggagggc tttggccctt gatgtatgg agccggaaaca
2161 agcagtggac ctgacattgc ccaagggttga agcaatgaat aaaagactgg gctccttgg
2221 ggatgagttt aaggagctt tttaccacc agattacaat cctgaaggaa aagttacca
2281 gagaaaaacac gataatgaag gttctggaaag caaaaggccc aaggtggagt attcagaaga
2341 ggagctgaag acccacaatca gcaagggttac gctggcaag ttcactgtgc ccatgtctaa
2401 agaggccctgc cgggcttacg ggctgaagag tggctgttgc aagcaggagc tgctggaaagc
2461 cctcaccatca cacttccagg actgaccaga ggcggccgtt ccagctggcc ttccgcgtt
2521 tggccaggctt gcttgcctt gtcctcaggcc agttaaatgc tttttcttctt gagctggaa
2581 gagtctaccc gacataagtc gagggactttt atgttttgc ggtttctgt tgccatgggt
2641 atggtgcgtac cctccactt tgctgttccat tactttactg cctgaataaaa gagccctaaag
2701 tttgtactaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa
  
```

//

SEQ ID 7

LOCUS AF093759 2493 bp DNA linear PRI 14-AUG-2001
 DEFINITION Homo sapiens heat shock protein HSP70 (HSPA7) gene, complete sequence.
 ACCESSION AF093759
 VERSION AF093759.1 GI:4139180
 KEYWORDS
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2493)
 AUTHORS Parsian,A.J., Sheren,J.E., Tao,T.Y., Goswami,P.C., Malyapa,R., Van Rheeden,R., Watson,M.S. and Hunt,C.R.
 TITLE The human Hsp70B gene at the HSPA7 locus of chromosome 1 is transcribed but non-functional
 JOURNAL Biochim. Biophys. Acta 1494 (1-2), 201-205 (2000)
 MEDLINE 20525459
 PUBMED 11072087
 REFERENCE 2 (bases 1 to 2493)
 AUTHORS Hunt,C.R., Malyapa,R., Parsian,A.J., Goswami,P.C., Van Rheeden,R. and Watson,M.S.
 TITLE Direct Submission
 JOURNAL Submitted (22-SEP-1998) Radiation Oncology, Washington University School of Medicine, 4511 Forest Park Blvd., St. Louis, MO 63108, USA
 FEATURES Location/Qualifiers
 source 1..2493
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q21.3"
 gene 233..2149
 /gene="HSPA7"
 /note="heat-inducible heat shock protein HSP70; possible pseudogene"
 BASE COUNT 554 a 676 c 794 g 469 t
 ORIGIN

SEQ ID 7

1 cccggggcggg cgggcgggag gctctcgact gggcgaaaag gtgcgggaag gttcgccggcg
 61 gcgggggtcgg ggaggtgcaa aaggatgaaa agcccgtgga agcggagctg agcagatccg
 121 agccgggctg gcggcagaga aaccgcaggg agagcctcac tgctgagcgc ccctcgacgg
 181 cggagcggca gcagccctcg tggcctccag catccgacaa gaagcttcag ccatgcaggg
 241 cccacgggag ctcgcgggtgg gcatcgaccc gggcaccacc tactcgtgcg tggcgtgttt
 301 tcagcagggc cgcgtggaga tcctggccaa cgaccaggc aaccgcacca. cggccagcta
 361 cgtggccttc accgacaccg agcggctggt cgggacgcgc gccaagagcc agggggccct
 421 gaaccccccac aacaccgtgt tcgatgcca gcggtgtatc gggcgaagt tcgcggacac
 481 cacgggtcag tcggacatga agcaactggcc cttccaggtg gtgagcggagg gcgcaagcc
 541 caaggtgcgc gtatgttacc gcggggagga caagacgttc taccggagg agatctcg
 601 catgggtctg agcaagatga aggagacggc cgaggcgtac ctggccagc ccgtgaagca
 661 cgcagtgtatc accgtgcccc cctatattcag taactcgac cgcaggcaca ccaaggacgc
 721 gggggccatc gcggggctca aggtgctgcc gatcatcaat gaggccacgg cagcagccat
 781 cgcctatggg ctggaccggc ggggcgcggg aaagcgcac ggtctcattt ttgacctggg
 841 tggggccacc ttcgatgtgt cgggtctctc cattgacgccc ggtgtcttt aggtgaaagc
 901 cactgctgga gataccacc tgggaggaga ggacttcgac aaccggctcg tgaaccactt
 961 catggaagaa ttccggcggaa agcatggaa ggacctgagc gggacaaga gtcgcctgc
 1021 caggctgcgc acagcctgtg agcgcgcac ggcgcaccccg tcctccagca cccaggccac
 1081 cctggagata gactccctgt tcgagggcgt ggacttctac aagtccatca ctcgtcccg
 1141 ctttgggaa ctgtgctcag acctcttccg cagcaccctg gagccgggtgg agaaggccct
 1201 gcgggatgcc aagctggaca aggcccagat tcatgacttc gtcctggggg gagggctcca
 1261 ctcgcattccc caaggtgcag aagttgctgc aggacttctt caacggcaag gagctgaaca
 1321 agagcatcaa ccctgatgag gctgtggcct atgggtctgc tgtgcaggcg gccgtgtga

1381 tgggggacaa atgtgagaaa gtgcaggata tcctgctgct ggatgtggct cccctgtctc
1441 tggggctgga gacagcaggat ggggtgatga ccacgctgat ccagaggaac gccactatcc
1501 ccaccaagca gacccagact ttcaccaccc actcgacaa ccagcctggg gtcttcatcc
1561 aggtgtatga ggttgagagg gccatgacca aggacaacaa cctgctgggg cgtttgaac
1621 tcattggcat ccctcctgccc ccacatggag tccccccagat agaggtgacg tttgacattg
1681 atgtaatgg catcctgagc gtgacagcca ctgacaggag cacaggtaaag gctaacaaga
1741 tcaccaatga caagggccgg ctgagcaagg aggaggtgga gaggatgggt catgaagccg
1801 agcagtacgg ggctgaggat gaggcccaga gggacagagt ggctgccaaa aactcgctgg
1861 aggcccattgt cttccatgtg aaaggttctt tgcaagagga aagccttagg gacaagattc
1921 ccgaagagga caggcgcaaa gtcaagaca agtgcagga agtccttgcc tggctggagc
1981 acaaccagct ggcagagaag gaggagtagt agcatcagaa gagggagctg gagcaaatct
2041 gtcgccccat cttctccagg ctctatgggg ggcctgggt ccctggggc agcagttgt
2101 ggcgtcaagc ccaccagggg gacccagca ccggcccccatttggagg gttgattgaa
2161 tggcccttcg tgataagtca gctgtgactg tcaggctat gctatggggc ttctagactg
2221 tcttctatga tcctgcctt cagagatgaa gggcttgggg gggcttccc tccaaagcta
2281 gaactttctt tccaggataa ctgaagtctt ttgactttt gggggggagg cggttcattcc
2341 tcttctgctt caaataaaaa gtcattaatt tattaaaact tttgtggcac tttaacattg
2401 ctttcaccta tattttgtgt atttgttac ttgtatgtat gaattttgtt atgtaaaata
2461 tagttataga cctaaataaa cttttaaaac tcc

//